SCIENCE CHINA
Chemistry

• REVIEWS •
• SPECIAL TOPIC • Frontiers in Polymer Science

December 2016 Vol.59 No.12: 1531–1540
doi: 10.1007/s11426-016-0231-3

Protein self-assembly: technology and strategy

Linlu Zhao†, Shanpeng Qiao† & Junqiu Liu*

State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun 130012, China

Received May 16, 2016; accepted July 3, 2016; published online October 27, 2016

Proteins, as the premier building blocks in nature, exhibit extraordinary ability in life activities during which process proteins mostly self-assemble into large complexes to exert prominent functions. Inspired by this, recent chemical and biological studies mainly focus on supramolecular self-assembly of proteins into high ordered architectures, especially the assembly strategy to unravel the formation and function of protein nanostructures. In this review, we summarize the progress made in the engineering of supramolecular protein architectures according to the strategies used to control the orientation and the order of the assembly process. Furthermore, potential applications in biomedical areas of the supramolecular protein nanostructures will also be reviewed.

protein assembly, supramolecular strategy, self-assembly, electrostatic interaction, metal-ligand coordination

Citation: Zhao L, Qiao S, Liu J. Protein self-assembly: technology and strategy. *Sci China Chem*, 2016, 59: 1531−1540, doi: 10.1007/s11426-016-0231-3

1 Introduction

Proteins are the premier building blocks in nature for the construction of various complicated architectures that perform their biological functions [1–4]. Majority of the characterized proteins are organized by forming oligomeric assemblies, which exhibits various advantages over the proteins alone, including stability, allosteric regulation, and functional control, etc. Especially, these protein assemblies, at the basis of numerous biological machines, can exert prominent functions in cells when hetero-monomeric forms were linked together to maintain life existence. Inspired by the exquisite biological structures, there has long been a desire to investigate protein self-assembly behavior, particularly the strategy to induce protein-protein association [5,6]. It is important to understand the assembly mechanism as self-assembly has become an efficient and promising strategy to fabricate a series of novel nanostructures [7]. For protein system, supramolecular self-assembly seems to be a powerful approach to construct complex architectures by utilizing non-covalent interactions, which provides the possibility to design and mimic the performance of native proteins [8,9].

In the past couple of years, there has been a growing interest in fabricating various hierarchical protein nanostructures [10–12]. However, the construction of protein-based assemblies with spatial order comparable to that of natural assemblies is quite challenging due to the complexity of protein interactions and structures. Therefore, the rational utilization of protein-protein interactions (PPI) may greatly improve the possibility of realizing protein self-assembly by amplifying the binding strength of protein association [13]. Fortunately, the increasingly sophisticated computational tools facilitate the manipulation of assembly process, which significantly promotes the development of novel nanostructures and biomaterials [14,15]. Moreover, the controllable interactions between proteins and metal ions, small molecules, or other proteins has led to approaches such as metal-ligand coordination, host-guest interactions, and electrostat-

†These authors contributed equally to this work.
*Corresponding author (email: junqiuliu@jlu.edu.cn)

ic interactions capable of engineering multiple assemblies. A variety of protein assemblies with interesting one-, two-, and three-dimensional (1D, 2D, 3D) morphologies and excellent biological functions are thus formed based on the above mentioned strategies. Protein-based assemblies have attracted considerable attention of contemporary researchers not only for their innate bioactivity, but also for realizing controllable arrangements at the molecular level with widespread applications in drug delivery, tissue engineering, light harvesting, and biomineralization [16–18].

The exploration to obtain *de novo* protein assemblies with superior properties has long been attempted and a large amount of achievements have been reported to date [19–21]. In this review, we summarize the progress made in the engineering of supramolecular protein architectures according to the strategies used to control the orientation and order of the assembly process. In addition, the potential applications in biomedical areas of the supramolecular protein nanostructures will also be reviewed.

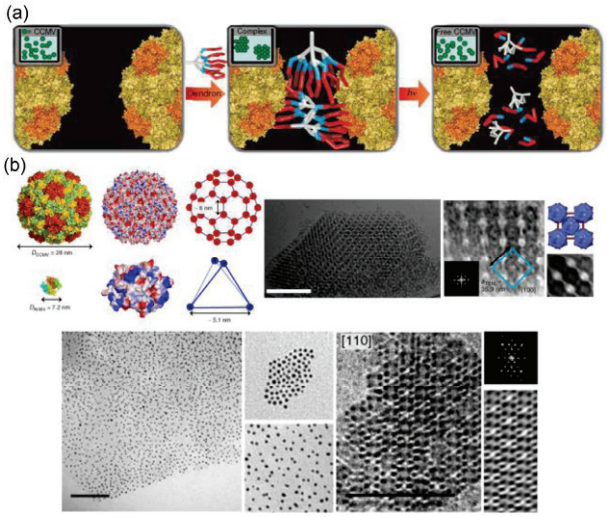

Figure 1 (a) Self-assembly and photo-triggered disassembly of CCMV-dendron complexes via electrostatic interactions; (b) three dimensional superlattices constructed by binary CCMV and biotin co-assembly [26,27] (color online).

2 Protein self-assembly based on supramolecular strategies

2.1 Electrostatic interactions induced protein self-assembly

Electrostatic interactions act as a significant role in organizing natural proteins into higher hierarchical architectures and at the same time stabilizing protein nanostructures [22–24]. Various amino acids with different charges distribute among the proteins, which lead to the proteins highly charged. Furthermore, as each protein possesses its individual isoelectric point (pI), proteins can present different charges under various pH values. What really makes sense to us is the charges distributed over protein surfaces, which can determine the assembly orientation provided that the surface circumstance is well exhibited [25]. With the development of computer simulating technology, protein-based self-assembly via electrostatic interactions has received rapid progress as the protein surface is better known. In addition, the long-distance weak electrostatic interactions can be reversibly regulated by altering ion strength of the solution, thus to control the assembly process.

2.1.1 Cage-like proteins formed crystal nanostructures

Kostiainen *et al*. [26] reported electrostatic induced protein self-assembly into well-defined micrometer-sized architectures. They utilized a type of negatively charged protein cages, icosahedral cowpea chlorotic mottle virus (CCMV), to achieve self-assembly and photo-triggered disassembly under the guidance of polyamine functionalized dendrons (Figure 1(a)). The branched cationic dendrons that bind to the negatively charged protein surface exhibited different binding affinity due to the effect of multivalency among various generations. It has been validated that the higher generation dendrons possess the potential to strongly bind the protein cage even at high salt concentration. Furthermore, transmission electron microscope (TEM) images showed micrometre-sized complexes with apparent dense hexagonal packing of the individual virus particles, which confirmed the multivalently electrostatic CCMV-dendron assembly. Upon exposing the complexes to ultraviolet (UV) light, photocleavage occurred with multivalent supramolecular interactions destroyed to release the free CCMV. As the CCMV has the spatially confined interior to encapsulate compounds and materials, this optically triggered disassembly can effectively regain the individual CCMV and further release the inclusions under pH control due to the innate pH-dependent assembly/disassembly feature of the virus. Significantly, this versatile method was also feasible to other negative charged protein cages. Magnetoferritin (MF), with an apo-ferritin cage to host synthetic superparamagnetic magnetite/maghemite (Fe_3O_4-g-Fe_2O_3) particles, was demonstrated to assemble with the dendrons based on the same electrostatic principles and tuned in the same manner. In view of the excellent features of these protein cages and the photocleavage disassembly at selected locations, this work held great potential for practical applications in targeted drug delivery and release.

Recently, the same group successfully constructed three-dimensional (3D) heterogeneous crystals of different protein-based building blocks alone via electrostatic interactions [27]. Binary and ternary crystals have always been significant mainly owing to their ability to integrate various components to yield multifunctional metamaterials [28]. Especially, the preparation of biologically active protein crystals held promise to simulate natural protein complexes and thus drove the development of exploring human diseas-

es. In this work, oppositely charged proteins, CCMV and avidin, were utilized to electrostatically assemble into heterogeneous crystals (Figure 1(b)). Avidin is a tetrameric glycoprotein with four positive patches distributed in the corner points of the tetrahedral geometry. Previous studies have demonstrated that a patchy interaction had great influence on the ultimate self-assembled nanostructures. Compared with the globular smoothly charged dendrimer who self-assembled with CCMV to obtain a face-centered cubic (fcc) crystal structure, the nanostructures assembled by CCMV-avidin were proved to be body-centered cubic (bcc) arrangement, where the CCMV was held together by the avidin in a non-close packing mode. According to the small-angle X-ray scattering (SAXS) and cryo-TEM characterization, the bcc crystal structure obtained by CCMV-biotin assembly can be clearly verified. They also found that proper concentration of electrolyte can promote the system to assemble into energetically favored crystals, however, excessive NaCl may lead to the charge screening and thus the crystals broken apart. Furthermore, avidin-biotin binding with extreme selectivity and affinity allowed the crystals to be readily functionalized. Biotin-tagged functional units such as 5-fluorescein, horseradish peroxidase (HRP) and AuNPs have been successfully integrated with CCMV-avidin crystals in pre- or post-ways to achieve functionalization without affecting their bcc structure. This work, introducing a third component into a binary protein crystal, paved way for the exploration of naturally complex protein assemblies.

2.1.2 Functional nanowires constructed by cricoid proteins

Apart from the protein superlattices, Liu's group [29] sought to utilize electrostatic self-assembly to construct high-ordered protein nanowires with prominent function. To achieve this goal, they chose stable protein one (SP1), a ring-like protein comprised of 12 subunits tightly bound to each other via hydrophobic interactions, as building block. The crystal structure of SP1 showed that it had high symmetry and negatively charged surface, which seemed to be an ideal model for electrostatic assembly. Positively charged semiconductor quantum dots (QDs) acted as an effective linker to induce SP1 ring into significantly ordered self-assembled nanostructures via spatial complement and electrostatic interactions. Intriguingly, QDs with different sizes (QD1 ~3.3 nm, QD2 ~5.8 nm, QD3 ~11.5 nm) can guide the protein into various superstructures including nanowires, clustered nanostructures and cross-linked network structures. dynamic light scattering (DLS), atomic force microscope (AFM) and TEM characterizations uniformly verified these morphologies well, which strongly proved the electrostatic self-assembly between oppositely charged SP1 and QDs and the size influence of the QDs to the assembly topography. Due to the capability of the QDs to harvest visible light energy, these assemblies where the QDs were separated in ordered arrangement to avoid self-quenching can serve as an ideal scaffold for a light-harvesting antenna as the energy can be transferred successively to adjacent QDs. In order to achieve efficient energy transfer, QD1 and QD3 were chosen, for the less overlap of the emission peaks compared with other pairs of QDs, to co-assemble with SP1 ring. Notably, the co-assembly nanowires exhibited a pronounced fluorescence resonance energy transfer (FRET) phenomenon based on QD1–QD3 as donor-acceptor pair, which can be illustrated from a series of fluorescent data (Figure 2(a–d)). AFM images showed two types of assembly mode corresponding to the direct energy transfer or multiple degradable donor-donor events and then energy transfer, respectively, and the latter form constructed a complete set of a light-harvesting system.

Utilizing the same concept, beyond "rigid" QDs, Liu *et al*. [30] further explored the possibility of "soft nanoparticles" to achieve assembly into functional nanostructures. Dendrimers, a type of uniform distributed organic macromolecules, can be accurately controlled of their sizes only by regulating their generations. Poly(amino amine) (PAMAM) dendrimers possessing multiple positive charges were regarded as the ideal "soft nanoparticles" to induce self-assembly with the cricoid SP1 as the SP1-QDs assembly behavior based on electrostatic interactions. Computer simulation shows that the fifth generation PAMAM dendrimer (PD5) had commendable size matching to guide SP1 assembly into linear architectures. As expected, the negatively charged SP1 and the positively charged PD5 packed

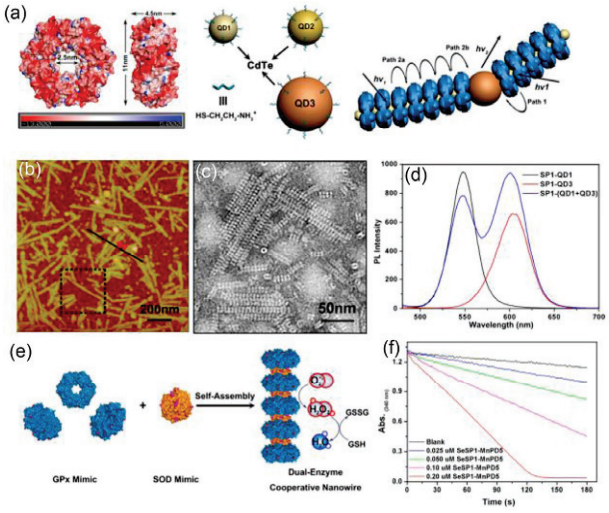

Figure 2 (a) Schematic representation of cricoid protein SP1, three kinds of CdTe QDs, and the self-assembled SP1-QD nanowires to mimic light harvesting system; (b) AFM images of the self-assembled nanostructures; (c) TEM analysis of SP1-QD assembly; (d) PL spectra of donor QD1 with SP1 (black line) and acceptor QD3 with SP1 (red line) in aqueous solutions and of their mixture solution (blue line); (e) design of dual-enzyme cooperative antioxidative system with both GPx and SOD activities; (f) plots of absorbance versus time during the catalytic reduction of H_2O_2 (0.5 mM) by GSH (1.0 mM) with 0 μM (black), 0.025 μM (blue), 0.050 μM (green), 0.10 μM (magenta), and 0.20 μM (red) SeSP1-MnPD5 at pH 7.0 and 37 °C [29,30] (color online).

orderly to fabricate defined nanowires, which can be fully validated by AFM and TEM. Moving forward, this binary nanostructure can realize its function as a dual-enzyme cooperative antioxidative nanozyme by introducing glutathione peroxidase (GPx) and superoxide dismutase (SOD) centers to SP1 and PD5, respectively (Figure 2(e, f)). They utilized genetic engineering approaches to obtain Se-SP1 endowed with excellent GPx activity. SOD catalytic center, manganese porphyrin, is modified on the surface of PD5 (MnPD5). Structure characterization showed that these two building block models with relevant catalytic active centers can also successfully assemble into high ordered nanowires. Furthermore, the assemblies exhibited both prominent GPx and SOD activities in protecting organisms especially the mitochondria from being oxidative damaged by effectively scavenging H_2O_2 and $O_2^{-\bullet}$ in the system. Significantly, this synergistic nanozyme possessed low cytotoxicity and good compatibility with human cells, which greatly improved the possibility of its practical applications in catalysis, pharmaceuticals and biosensors.

2.2 Metal coordination-driven protein self-assembly

The formation of functional protein and protein complexes benefited from the protein-metal coordination is ubiquitous in nature. The transition metals such as Fe, Cu and Zn contribute to the function and structure of protein in nature, not only because their redox and high Lewis acidities as catalytic center ensure the indispensable bio-processes, but also their high affinity of combination with several amino acids (such as His, Cys, Asp, and Glu which show as donor ligands to form appreciably strong coordination with metal) enables the critical folding and interactions between protein complexes. In addition, owing to the inherent symmetry and directionality of individual metal, the stability, and the stimuli-responsiveness (chelators, pH, and redox station), metal coordination has become a reliable strategy for driving the redesigned and engineered proteins to form high-order, functional and smart protein self-assemblies.

It was the two decades that the metal coordination is employed on protein aggregation and becomes an efficient method for designing the protein self-assembly. Many efforts were focused on construction of protein oligomers via coordination chemistry, and there were mainly three strategies to achieve this goal: modification with the non-natural metal chelators, design several metal chelate motifs on the surface of proteins, and co-assembly with metal coordinative agent.

2.2.1 Modification with non-natural metal chelators

The method that covalent modification of the non-natural metal chelation seems a straight way. Jensen *et al.* [31] has reported a trimer protein complexes building up from 2,2′-bipyridine (Bipy)-derivatized insulin monomers upon Fe^{2+} coordination. Tezcan and coworkers also describe an engineered protein (MBPPhen2, a cyt cb_{562} variants modified with 1,10′-phenanthroline on surface) to yield dimer and further D_2-symmetric tetramer in the presence of Zn and Ni metal ions [32]. Based on this strategy, one-dimensional bundles [33] and 3D polymer or porous structures [32,34] have been generated successfully (Figure 3). However, because of the flexibility of metal chelate and few anchored points, the protein oligomers show less ordered or only present under harsh condition while self-assembling.

2.2.2 Designing metal motifs by site-directed mutagenesis

In the other approach, well-defined protein architectures could also be obtained via installing bi/multidentate motifs on the helical surface to increase rigidity and confirm the orientation. Tezcan *et al.* [35] first employed the strategy on cyt cb562 (a four-helix bundle hemeprotein) and fine designed metal coordination motifs on its surface for protein self-assembling through genetic engineering. The redesigned protein (MBPC1) with two i/i+4 bis-histidine motifs (His59/His63 and His73/His77) has been shown to self-assemble and generate the oligomers which have been characterized by DLS, nuclear magnetic resonance (NMR), sedimentation velocity (SV) and crystallographic in the presence of Zn^{2+}, Cu^{2+}, and Ni^{2+}. All experiments above clearly demonstrate the formation mechanism of the metal-mediated oligomers, that is, self-assembling is going ahead with the metal-protein coordination and assistant protein-protein interaction. Noteworthily, different topologies and states of aggregates can be obtained upon distinctive metal ions, of which stereochemical preferences tight control binding orientations [36]. For example, the tetrahedral Zn^{2+} induce cyt cb_{562} variant to form D_2-symmetrical tetramer whereas it is demonstrated that only C_2-symmetric dimer and C_3-symmetric trimer are achieved upon Cu^{2+} and Ni^{2+} coordination, respectively, due to the different steric

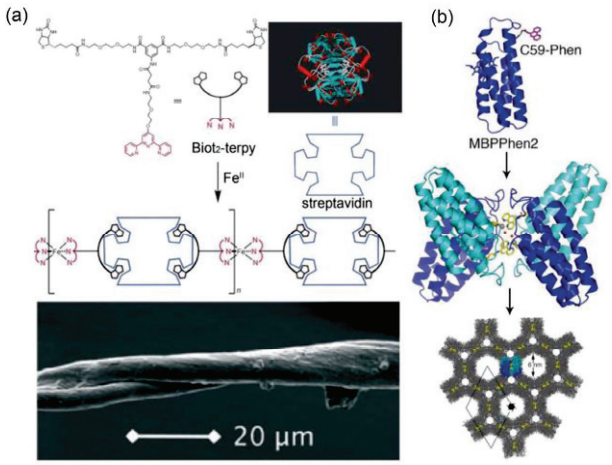

Figure 3 (a) Fe-directed streptavidin liner assembly bundles using 2,2′:6′,2′-terpyridine (Terpy)-biotin$_2$ as connector and metal chelate [33]; (b) Zn-mediated crystal lattice. Tetramer (Zn_2:MPBPhen2_4) complex is formed and further crystalize to 3D porous protein frameworks upon Zn metal coordination. The highlight Phen groups make great contribution to the generation of protein tetramer complex [32] (color online).

arrangement of metal-coordination motifs on surface and the unique protein interface interactions. Using this cyt cb_{562} variant as a starting point, Tezcan *et al.* [37–39] engineered complementary interfaces of protein through metal templated interface redesign strategy (MeTIR), a computational modeling of sidechain interaction around metal motifs to stable the metal coordinate complexes, and achieved a stable tetrameric oligomer with selective Zn metal ion. Meanwhile, another cyt cb562 derivative was also subsequent redesigned around the protein-protein interface by MeTIR with ten mutations (Lys27Glu, Asp28Lys, Thr31Glu, Arg34Leu, Leu38Ala, Asn41Leu, His59Arg, Asp66Ala, Val69Met and Leu76Ala) and finally self-assembled into highly ordered periodic protein arrays and nanotubes (Figure 4) of large size in the presence of Zn^{2+} [40]. X-ray crystallography and cryoelectron microscopy (cryo-EM) have clearly demonstrated the principle and structural details. A C_2-symmetric dimer building block was first produced and stabilized via unsaturated Zn^{2+} coordination (His73/His77 from one and His63 from another) and hydrophobic or polar interactions of the redesigned surface respectively. The self-assembling was further induced upon coordination of Zn^{2+} sites (His73/His77/His63 and Ala1 or Asp39) and formed the protein architectures. Noteworthily, the 2D planar structure (pH 5.5) and helical nanotubes (pH 8.5) could

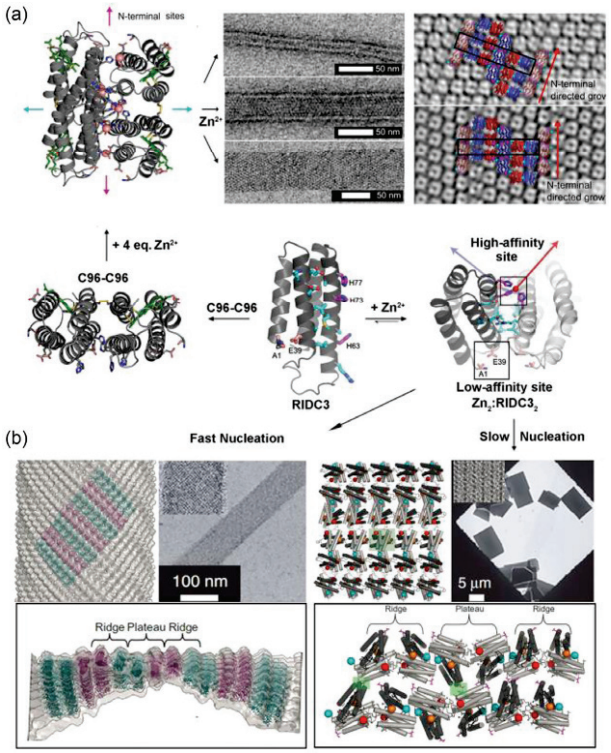

Figure 4 (a) The Zn metal induced D_2 symmetric tetramer is generated from disulfide-linked protein dimers and act as building block of larger protein nanotubes obtained upon further Zn coordination [38]; (b) a C_2-symmetric dimer (Zn_2:$RIDC3_2$) is produced through redesigned face and metal coordination. Nanotubes and 2D arrays are achieved under fast (pH 8.5 or high [Zn]/[protein] ratios) and slow (pH 5.5 or low [Zn]/[protein] ratios) nucleation conditions respectively [40] (color online).

be achieved and transformed under different pH and [Zn]:[protein] ratios [40,41].

This strategy has also shown viable in more complicated protein system. Liu and colleagues [42] have fused metal chelate motifs (a His_6-tag) on the N-terminal of Sj-GST, a C_2-symmetric homodimer glutathione S-transferase from Schistosoma japonicum with globular profile, to generate a liner functional protein nanowire. Two His_6-tags on the opposite sides of protein surface endow engineered protein liner symmetry and less influence on enzymatic center in the middle of protein upon Ni^{2+} coordination. AFM, circular dichroism (CD) and enzymatic activity assays provide the design very well. DSC and native-polyacrylamide gel electrophoresis (PAGE) also demonstrate the Ni^{2+} induced reversible assembly/disassembly process (Figure 5(a)). Furthermore, another Sj-GST variant with V shaped chelating agents was generated via site-directed mutation of bis-His motif (His137/His138) on the alpha helix location of Sj-GST monomer. Combining the asymmetric metal coordination of bis-His motifs and secondary protein interaction, protein nanorings have been achieved through precisely controlling the growth direction and salt-concentration accommodated surface electrostatic stabilization [43]. DLS, AFM, negative-stain TEM measurements and theoretical calculations result into tunable nanorings with diameters ranging from 96 to 367 nm at different salt concentration, which provides metal coordination a promising method for

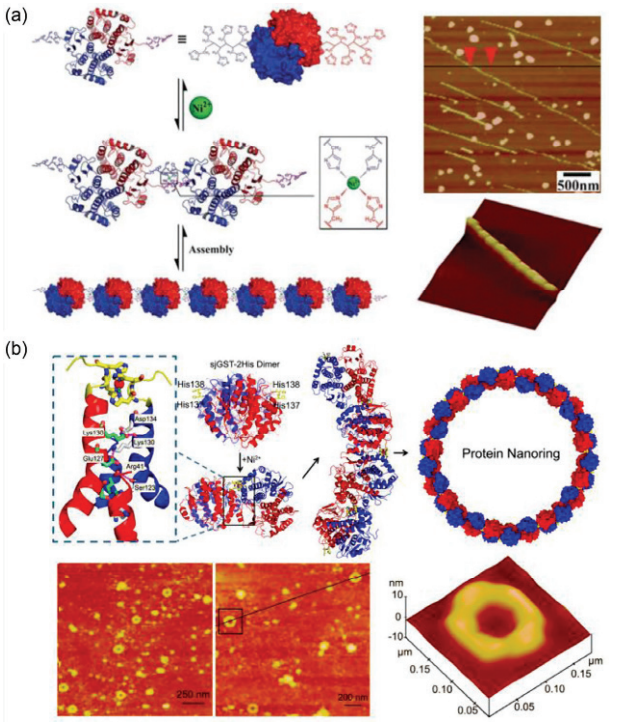

Figure 5 Self-assembled linear and nanoring from engineered hdGST via metal coordination. (a) Linear protein assembly [42]; (b) metal-mediated protein nanoring, the orientation of motifs is designed with a shift from symmetric axis to regulate the geometry of protein assembly [43] (color online).

construction of elaborate protein architectures (Figure 5(b)).

2.2.3 Protein crystals induced by metal-organic linker-directed interactions

Inspired by the metal-organic frameworks (MOFs) and crystallography, protein co-assembling with metal-organic linker has become a novel and advanced methodology in metal induced protein 3D assembly. Not only because the metal-anchored protein complexes provide more symmetries which can be precisely designed via controlling location of metal anchor sites on protein surface to coordinate with organic linkers, but also the 3D crystals could be obtained within a designed stacking manner and under a less harsh condition with the aid of this strategy. Tezcan *et al.* [44] pioneered this idea in designing 3D protein crystals via Zn^{2+} coordination, which includes the symmetric polytopic organic linkers and highly stable human ferritin, an octahedral (432) symmetric spherical protein as a scaffold. By site-directed mutation of histidine at 122nd amino acid which is located on the C_3-symmetric pores and exposed to exterior surface, a tetrahedral geometric metal-protein node was constructed owing to the coordination of tripodal His-motifs ferritin variant (^{T122H}ferritin) with Zn metal ions. Upon addition of Zn^{2+}, the 3D crystal was further obtained through co-assembling of ^{T122H}ferritin and benzene-1,4-dihydroxamic acid (H_2bdh), a ditopic organic linker with high Zn^{2+} binding affinity. Compared with Zn-^{T122H}ferritin crystals in a face-centered cubic (fcc) lattice via vapor diffusion method, the self-assembled bdh-Zn-^{T122H}ferritin crystals demonstrate a body-centered cubic (bcc) arrangement in lattice by design and are generated in a wide range of solution conditions (pH 5.5–9.5) (Figure 6). X-ray diffraction, light micrograph, DLS measurements are used to sort the appropriate organic linker, displaying the arrangement in Zn-^{T122H}ferritin and bdh-Zn-^{T122H}ferritin crystals and coordination process. This strategy provides a further methodology in constructing stable functional materials.

2.3 Protein assembly based on host-guest interactions

Host-guest interactions, small-molecule based driving forces with high affinity, selectivity and reversibility, have exhibited great potential in the construction of desired protein nanostructures [45]. The host molecule can strongly capture the guest units into its cavity, thus to direct the recognition and assembly of proteins. Brunsveld *et al.* [46–49] have demonstrated the feasibility of utilizing host-guest interactions to achieve protein dimerization in the previous work. Macrocycles seem to be the ideal scaffolds as host molecules due to their capability to afford a wide range of recognition motifs in aqueous media. Host/guest molecules are alternatively modified onto protein surfaces to allow manipulating the protein assemblies by association of host-guest pairs [50]. Up to now, host-guest pairs including cucurbit[8]uril (CB[8])&FGG, CB[8]&methylviologen&alkoxynaphthalene and β-cyclodextrin (β-CD)&lithocholic acid motifs are validated to have strong binding forces to induce protein association. Owing to the reversibility of the host-guest interactions, protein assembly/disassembly process can be readily controlled, which would in favor of enzyme activation by forming protein dimer and controlling over enzyme activity.

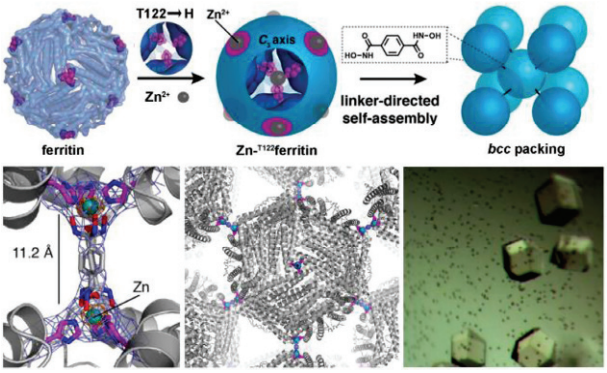

Figure 6 Metal/linker-directed 3D crystal of engineered ferritin. The C_3 pores are designed as metal binding sites via T122H mutation. The bdh-Zn-^{T122H}ferritin crystals with bcc type lattice are generated through co-assembling of ^{T122H}ferritin and benzene-1,4-dihydroxamic acid (H_2bdh) [44] (color online).

Liu *et al.* [51] successfully extended host-guest induced protein association from previous dimer to higher hierarchical nanostructures. They selected a C_2-symmetric glutathione S-transferase (GST) homodimer with tripeptide FGG fused onto its N-termini (GST-FGG) as building block model. Previous knowledge have demonstrated that FGG had strong binding force with CB[8] at the ratio of 2:1, which can serve as an effective host-guest based inducer to guide protein association. Inspired by this research, Liu's group [51] constructed GST nanowires by utilizing CB[8] to combine the two FGG-tags protruding from the asymmetric N-termini of the homodimer in an opposite orientation (Figure 7). Isothermal titration calorimetry (ITC) data proved that CB[8] formed a ternary complex with FGG- GST and the binding constant of CB[8]-(FGG-GST)$_2$ was found to be 2.9×10^{12} M^{-2}, which indicated a high affinity of this host-guest association. Furthermore, this complex was also examined by size exclusion chromatography (SEC) to show a gradually increased peak intensity with the addition of CB[8] and simultaneously the reduced peak intensity of free FGG-GST, which validated that the specific host-guest interactions can form large and stable assemblies. Additionally, the detailed morphology had been characterized by AFM where uniform nanowires can be clearly observed with the height consistent with GST's theoretical size. Beyond the linear structures, the assemblies can also be well functionalized to become GPx mimics. With the aid of computer simulation, a cysteine auxotrophic expression system was chosen to incorporate catalytic selenocysteine

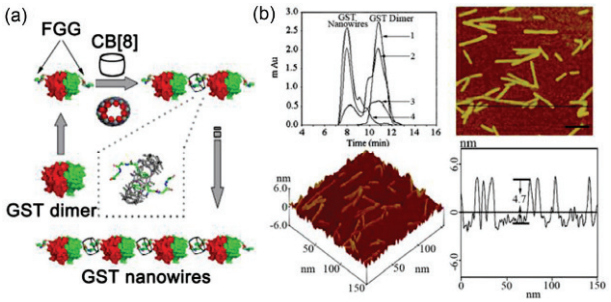

Figure 7 (a) Formation of protein nanowires by CB[8]-based host-guest interactions; (b) SEC of mixtures of FGG-GST and various concentrations of cucurbit[8]uril and AFM characterization of the nanowires [51] (color online).

into the active center, thus to possess antioxidative capacities. This nanozyme exerted excellent antioxidative properties in protecting mitochondria from oxidative stress, which may have promising potential for applications in biosensors, catalysis, and pharmaceuticals.

Recently, the same group again made a great development in host-guest based protein assembly [52]. Combined with fusion protein and supramolecular chemistry strategies, they successfully obtained calcium responsed spring-like nanowires (Figure 8). By chemical synthesis molecules or thermal stable DNA, the researchers successfully constructed supramolecular structure having a spring sports properties. Using protein as building block to design artificial nano-spring has not been reported yet. Here, a calcium-binding protein recoverin is designed to be an allosteric core. A phenyalanine-glycine-glycine tripeptide (FGG) is

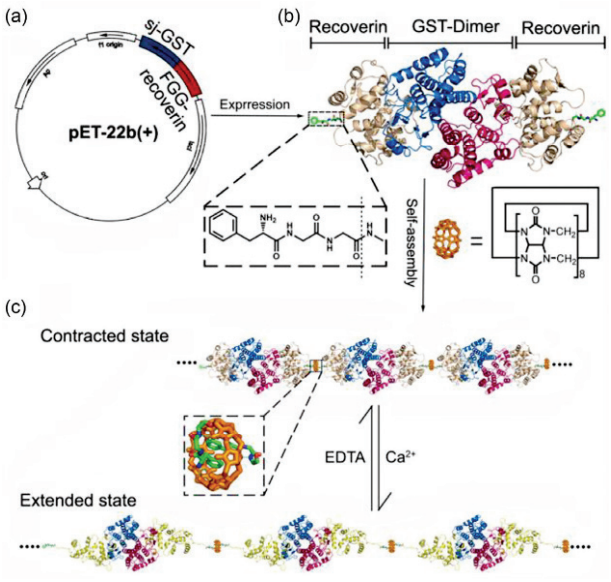

Figure 8 Design of the fusion protein FGG-recoverin-GST and protein nano-springs. (a) The recombinant plasmid of fusion protein FGG-recoverin-GST; (b) the detailed structure of the FGG-recoverin-GST dimer; (c) FGG-recoverin-GST assemblies induced by cucurbit[8]uril and the contracted state and the extended state of protein nano-spring controlled by binding or unbinding of Ca^{2+} [52] (color online).

fused at its N-terminus to form a 2:1 composite structure with CB[8] as host-guest recognition linker. At the same time dimeric enzyme glutathione S-transferase (GST) was fused on the coaxial alignment of C-terminus as oligo linker. The new recombinant protein combined stimulus response, supramolecular recognition and protein dimerization properties together. When adding an appropriate proportion of CB[8], this fusion protein self-assembled to tight liner nanostructures. The protein self-assembling structures were characterized by dynamic light scattering and atom force microscope both with and without Ca^{2+}. The protein self-assemblies were confirmed to be single nanowires of different lengths and flexibilities with Ca^{2+} "on-off" repeated stimuli. The details showed that the protein nanowires successfully integrated the structural change of each building block. Compared to the "contracted state" without the presence of Ca^{2+}, the length of the "extended state" with Ca^{2+} can raise as much as 50%. Besides, other experiments also confirmed that the regulation of ion response were stable and reversible. This protein assembly is a novel, sensitive nanospring that can be a preliminary simulation of natural elastic tissue. The study also provides a new idea to construct smart biological nanostructures.

2.4 Supramolecular protein assemblies constructed through protein-protein interaction

Protein-protein interaction owing to the stabilization of extensive and specific non-covalent force plays crucial roles in native protein assemblies, many indispensable bioprocesses such as signal transduction, catalysis, and immunity are based on the protein-protein interaction revealing rigorously specific recognition, high affinity, and enzymatic function which can be served as a powerful force to induce protein-protein recognition and aggregation.

2.4.1 Homo-fusion protein with specific symmetry

Among the strategies which are used on construction of well define protein assemblies, genetic fusion and swapping demonstrate fascination of designing fine one, two, and three-dimension protein arrangement via utilizing the initial symmetry and profile of protein subunits or motifs. It is a straightforward way to construct protein self-assembly by one-step method via fusing proteins with multiple binding subunits. Yeates *et al.* [53] genetically fused two naturally dimeric proteins and finally obtained the homology protein filament successfully. In the work, two homodimer proteins, carboxylesterase and influenza virus matrix protein M1, are connected by a five-residue helical rigid linker to generate a liner geometry engineer protein (Figure 9(b)). The two dimeric subunits tend to dimerize through protein-protein interaction individually and induce the formation of bipolar filament with head-to-head, tail-to-tail arrangement. TEM images discover linear filaments with about 4 nm wide in each strand and find further networks or bundles, which

demonstrate the successful construction of 1D protein assembly by protein fusion.

A more complicated protein cage was further reported through a symmetric fusion by controlling the relative orientation and expanding the respective symmetric geometry of distinct subunits [53–55]. The homo-trimeric bromoperoxidase and homo-dimeric M1 matrix protein of influenza virus is genetic fused by a nine-residue helical rigid linker which rotate the symmetric axes of the two subunits with an angle value of 109.5° [53]. The fusion protein with the specific geometry could self-assemble to a homogeneous 12-subunit cage with unique size after changing two residues via site-directed mutagenesis [54]. TEM and crystallography are applied in the protein cage to elucidate the precision of the symmetric fusion strategy (Figure 9(a, c)).

2.4.2 Construction protein assemblies via co-assem- bling of hetero-fusion protein

Different from the protein assembly generated in one step upon homology engineer protein, which often results in inclusion body (protein aggregates or precipitates formed through hydrophobic or ionic interactions or both, almost exclusively contain the over expressed protein) in expression system. Suzuki *et al.* [56] developed an indirective route to filament assembly by combining two kinds of "hetero-nanolego" components constructed via fusion of protein subunits, obtained the timing-controlled and size-regulated filament from solution. Each kind of the nanolego consist of a superoxide reductase (SOR), a structurally stable symmetrical homotetramer protein, as the structural elements and one of mouse PDZ domain (PDZ) or PDZ-binding peptide (Zpep), which two can associate with K_D (dissociation constant) of ~10^{-7}, as the binding elements. The final two kinds of nanolego, $(SOR\text{-}PDZ)_4$ and $(SOR\text{-}Zpep)_4$, is realized by fusing PDZ or Zpep at the C-terminal end of SOR without linker. By mixing $(SOR\text{-}PDZ)_4$ protein with $(SOR\text{-}Zpep)_4$ protein, the two kinds fused proteins can heterogenously self-assemble to $[(SOR\text{-}PDZ)_4(SOR\text{-}Zpep)_4]_n$ linear filament with the aid of PDZ-Zpep interaction of each nanolego. The morphologies of the protein assemblies are characterized by TEM and SEM. Surface plasmon resonance (SPR) assays were also employed to analysis kinetics process, indicating a high association constant of PDZ-Zpep interaction and hetero-assemble process of filament (Figure 10(a)).

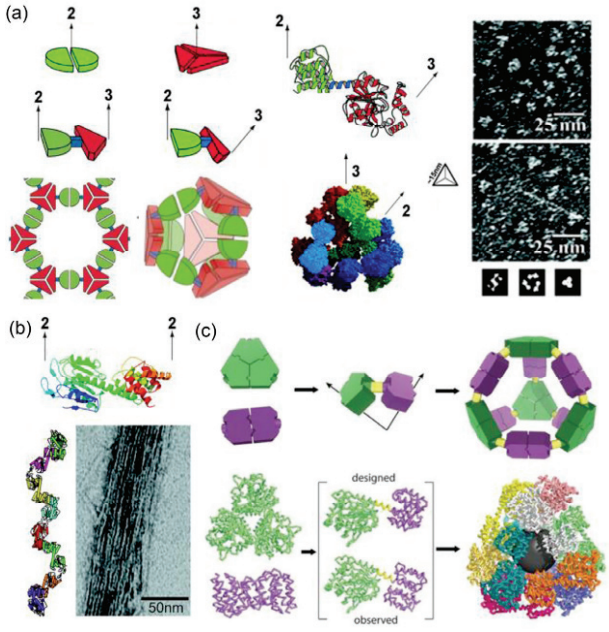

Figure 9 Filaments and cages constructed by genetically fusing two subunits of natural protein. Controlling the symmetry geometry and geometry of the symmetry axes, various regular protein assemblies can be obtained. (a) Protein cages; (b) protein filaments assembled from fusion protein with trimeric and dimeric domain [53]; (c) protein cage with unique size is produced with site-directed mutagenesis [54] (color online).

Appling the same strategy, Wei and colleagues [57] fused the octameric leucine dehydrogenase (LDH) and dimeric formate dehydrogenase (FDH) with PDZ domain and its ligand (PDZlig) respectively, reported a more complicated 2D bi-enzymatic networks with function of higher NAD(H) recycling efficiency. Same as the prior work we refer, the LDH and FDH domain in the fusion protein takes a part of structural element, while the PDZ and PDZlig respectively added to the C-termini of LDH and FDH domain is acted as binding element. The two-dimensional layer-like structures is successfully obtained as assumption because of the multiple symmetry afforded by the octameric LDH-PDZ which ensure extension of protein assembly through PDZ-PDZlig interaction and oligomerization of LDH, FDH subunits. SEM, AFM and epifluorescence exhibited successful generation of the 2D protein layer with large size *in vitro* and *in vivo*, enzymatic activity assays provided that the 2D protein networks seem an outstanding *L*-tert-leucine-synthesizer with higher NAD(H) recycling efficiency com

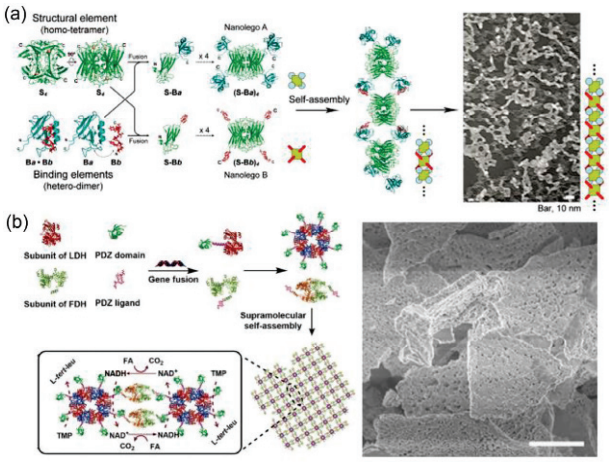

Figure 10 (a) Filaments obtained based on interaction of hetero-"nanolego" by genetically fusing superoxide reductase with PDZ domain and PDZ-binding peptide respectively [56]; (b) highly ordered self-assembly of oligomeric enzymes with higher NAD(H) recycling efficiency [57] (color online).

pared to the unassembled LDH and FDH. Therefore, fusion protein is a flexible, convenient and practical strategy for construction of self-assembly via genetic engineering (Figure 10(b)).

2.4.3 Artificial protein-protein interaction via computational redesign

Except fusing and swapping the natural proteins and utilizing intrinsic interactions between oligomeric subunits to induce the protein self-assembly, artificial protein-protein interaction realized through computational surface redesign is another advanced strategy for construction of highly ordered multidimensional protein architecture in the nowadays. Baker *et al.* [58] described several two-dimensional (2D) protein arrays mediated by computational redesigned noncovalent protein-protein interfaces. During the redesign process, symmetric docking was used to search the most shape-complementary interfaces between different oligomer copies, followed by sequence design calculations which can minimize the energy of well-packed interfaces. Monte Carlo searches, Rosetta all-atom energy, and a series of sequence design, optimization, were carried out by turn to design the shape complementary interfaces, hydrogen bond with low energy. Finally, 62 designs were picked out for experimental characterization using this strategy, and only four among of them shown soluble 2D layers lattices after expression. Negatively stained TEM, Fourier transformation, and electron cryomicroscopy (cryo-EM) have clearly illustrated the highly ordered 2D layers upon morphologies, electron diffraction, and projection map at high resolution, revealing that this strategy is a general approach to design protein assemblies.

In another example, 24-subunit, 13-nm diameter and 12-subunit, 11-nm diameter protein nanocage was designed with octahedral and tetrahedral symmetry respectively using engineered C_3-symmetric trimeric protein as building block [59]. The redesign was executed on symmetric docking, Rosetta Design calculations to optimize complementary interfaces, minimize binding energies, and select fit amino acids. Native PAGE and size-elution chromatograms (SEC) provided the nanostructures of several redesigned candidates and derivatives. The structural information of size-unique reconstructed O3-33 protein was characterized in detail by TEM and cryo-EM, revealing the actual assembly configuration and close match of redesign surface was realized through computational redesign (Figure 11).

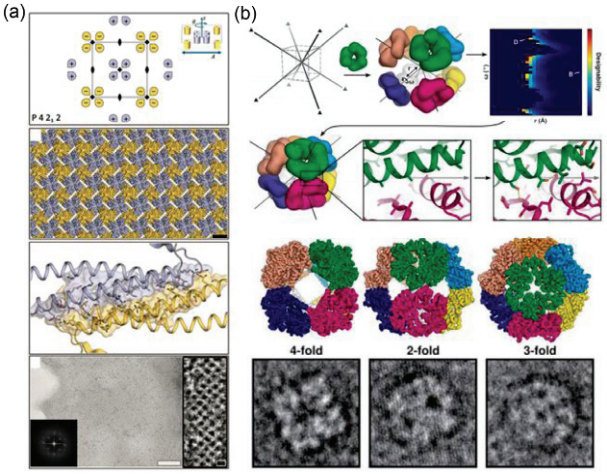

Figure 11 (a) Extended 2D protein arrays designed via general computational design strategy; (b) self-assembled protein cages directed by interfacial noncovalent networks and the crystal structure of designed O3–33 [58,59] (color online).

3 Conclusions and perspectives

Inspired by natural complex or based on chemical interactions, many interactions and strategies have been explored in the field of protein self-assembly recently, and also shown as powerful methodologies to induce protein self-assembly. Notably, the finial structures and functions of supramolecular protein self-assembly are not only determined by the various structures and intrinsic characteristics of protein building blocks but also the versatility of induced interaction. Thereby reveals infinite further applications of biomaterials self-assembled through supramolecular strategies.

Nowadays, protein self-assembly focuses on, in my opinion, the exploration of novel supramolecular strategies and functionalization of biomaterials. Giving the perspective of this field, several tendencies will be developed on the protein self-assembly area: (1) Novel methodologies and functionalization. Even through lots of interactions have been employed to induce protein self-assembly, novel strategies are still developed for designing protein superstructures in this potential area. Due to mostly functionalization of the protein self-assembly system is still preliminary to date, further efforts will be focused on the utilization of intrinsic functions of protein building blocks or bringing in the novel functions via chemical methods, reveals the protein supramolecular assemblies promising materials with infinite properties and versatile applications. (2) Construction of dynamic controlled protein self-assembly. The early controlled self-assembly of protein has been focused on the switch between assembling and disassembling, but the smarter assemblies with conformational changeability upon external stimuli still remain challenge and could attract more interests in future, therefore benefit more fascinating protein materials with dynamic controlled configurations and functions. (3) Self-assembling *in vivo*. the charm of protein self-assembly which make a splash root in the natural protein self-assembling in cell, for example, cytoskeleton, signal path and protein recruiting execute function and decide the destiny of cells. Taking advantages of biocompatibility and biodegradation, many cellular applications including drug delivery, vaccine design could be realized *in*

vivo with high efficiency and low cytotoxicity compared with self-assembly of organic molecules. The protein self-assembly *in vivo* thus demonstrate to be ascendant field and challenge to construct protein assemblies in the more complicated cellar environment than *in vitro*. In summary, this review provides a brief introduction on the protein self-assembly field and afford several extracted strategies for designing protein superstructures, hoping it is helpful to further exploration of this research aspect.

Acknowledgments This work was supported by the National Natural Science Foundation of China (21234004, 21420102007, 21574056, 91527302), and the Chang Jiang Scholars Program of China.

Conflict of interest The authors declare that they have no conflict of interest.

1 Jie KC, Zhou YJ, Yao Y, Huang FH. *Chem Soc Rev*, 2015, 44: 3568–3587
2 Pieters BJGE, van Eldijk MB, Nolte RJM, Mecinović J. *Chem Soc Rev*, 2016, 45: 24–39
3 Salgado EN, Radford RJ, Tezcan FA. *Acc Chem Res*, 2010, 43: 661–672
4 Yeates TO, Padilla JE. *Curr Opin Struct Biol*, 2002, 12: 464–470
5 King NP, Lai YT. *Curr Opin Struct Biol*, 2013, 23: 632–638
6 Zhang J, Zheng F, Grigoryan G. *Curr Opin Struct Biol*, 2014, 27: 79–86
7 Fegan A, White B, Carlson JCT, Wagner CR. *Chem Rev*, 2010, 110: 3315–3336
8 Luo Q, Dong ZY, Hou CX, Liu JQ. *Chem Commun*, 2014, 50: 9997–10007
9 Bai YS, Luo Q, Liu JQ. *Chem Soc Rev*, 2016, doi: 10.1039/C6CS00004E
10 Oohora K, Onoda A, Hayashi T. *Chem Commun*, 2012, 48: 11714–11726
11 Oohora K, Hayashi T. *Curr Opin Chem Biol*, 2014, 19: 154–161
12 Maeda Y, Matsui H. *Soft Matter*, 2012, 8: 7533–7544
13 Carlson JCT, Jena SS, Flenniken M, Chou T, Siegel RA, Wagner CR. *J Am Chem Soc*, 2006, 128: 7630–7638
14 Sakai F, Yang G, Weiss MS, Liu YJ, Chen GS, Jiang M. *Nat Commun*, 2014, 5: 4634
15 Yang G, Zhang X, Kochovski Z, Zhang YF, Dai B, Sakai F, Jiang L, Lu Y, Ballauff M, Li XM, Liu C, Chen GS, Jiang M. *J Am Chem Soc*, 2016, 138: 1932–1937
16 Biswas S, Kinbara K, Niwa T, Taguchi H, Ishii N, Watanabe S, Miyata K, Kataoka K, Aida T. *Nat Chem*, 2013, 5: 613–620
17 Li Q, So CR, Fegan A, Cody V, Sarikaya M, Vallera DA, Wagner CR. *J Am Chem Soc*, 2010, 132: 17247–17257
18 Suci PA, Kang S, Young M, Douglas T. *J Am Chem Soc*, 2009, 131: 9164–9165
19 Kitagishi H, Kakikura Y, Yamaguchi H, Oohora K, Harada A, Hayashi T. *Angew Chem Int Ed*, 2009, 48: 1271–1274
20 Onoda A, Kakikura Y, Uematsu T, Kuwabata S, Hayashi T. *Angew Chem Int Ed*, 2012, 124: 2682–2685
21 Oohora K, Burazerovic S, Onoda A, Wilson YM, Ward TR, Hayashi T. *Angew Chem Int Ed*, 2012, 51: 3818–3821
22 Pan XY, Yu SY, Yao P, Shao ZZ. *J Colloid Interf Sci*, 2007, 316: 405–412
23 Kostiainen MA, Ceci P, Fornara M, Hiekkataipale P, Kasyutich O, Nolte RJM, Cornelissen JJLM, Desautels RD, van Lierop J. *ACS Nano*, 2011, 5: 6394–6402
24 Kostiainen MA, Pietsch C, Hoogenboom R, Nolte RJM, Cornelissen JJLM. *Adv Funct Mater*, 2011, 21: 2012–2019
25 Sun HC, Zhang XY, Miao L, Zhao LL, Luo Q, Xu JY, Liu JQ. *ACS Nano*, 2016, 10: 421–428
26 Kostiainen MA, Kasyutich O, Cornelissen JJLM, Nolte RJM. *Nat Chem*, 2010, 2: 394–399
27 Liljeström V, Mikkilä J, Kostiainen MA. *Nat Commun*, 2014, 5: 4445
28 Kostiainen MA, Hiekkataipale P, Laiho A, Lemieux V, Seitsonen J, Ruokolainen J, Ceci P. *Nat Nanotechnol*, 2013, 8: 52–56
29 Miao L, Han JS, Zhang H, Zhao LL, Si CY, Zhang XY, Hou CX, Luo Q, Xu JY, Liu JQ. *ACS Nano*, 2014, 8: 3743–3751
30 Sun HC, Miao L, Li JX, Fu S, An G, Si CY, Dong ZY, Luo Q, Yu SJ, Xu JY, Liu JQ. *ACS Nano*, 2015, 9: 5461–5469
31 Munch HK, Heide ST, Christensen NJ, Hoeg-Jensen T, Thulstrup PW, Jensen KJ. *Chem Eur J*, 2011, 17: 7198–7204
32 Radford RJ, Lawrenz M, Nguyen PC, McCammon JA, Tezcan FA. *Chem Commun*, 2011 47: 313–315
33 Burazerovic S, Gradinaru J, Pierron J, Ward TR. *Angew Chem Int Ed*, 2007, 46: 5510–5514
34 Radford RJ, Tezcan FA. *J Am Chem Soc*, 2009, 131: 9136–9137
35 Salgado EN, Faraone-mennella J, Tezcan FA. *J Am Chem Soc*, 2007, 129: 13374–13375
36 Salgado EN, Lewis RA, Mossin S, Rheingold L, Tezcan FA. *Inorg Chem*, 2009, 48: 6082–6084
37 Salgado EN, Ambroggio XI, Brodin JD, Lewis, Richard A, Kuhlman, B, Tezcan FA. *Proc Natl Acad Sci USA*, 2010, 107: 1827–1832
38 Brodin JD, Medina-Morales A, Ni T, Salgado EN, Ambroggio XI, Tezcan FA. *J Am Chem Soc*, 2010, 132: 8610–8617
39 Brodin JD, Smith SJ, Carr JR, Tezcan FA. *J Am Chem Soc*, 2015, 137: 10468–10471
40 Brodin JD, Ambroggio XI, Tang C, Parent KN, Baker TS, Tezcan FA. *Nat Chem*, 2012, 4: 375–382
41 Brodin JD, Carr JR, Sontz PA, Tezcan FA. *Proc Natl Acad Sci USA*, 2014, 111: 2897–2902
42 Zhang W, Luo Q, Miao L, Hou CX, Bai YS, Dong ZY, Xu JY, Liu JQ. *Nanoscale*, 2012, 4: 5847–5851
43 Bai YS, Luo Q, Zhang W, Miao L, Xu JY, Li HB, Liu JQ. *J Am Chem Soc*, 2013, 135:10966–10969
44 Sontz PA, Bailey JB, Ahn S, Tezcan FA. *J Am Chem Soc*, 2015, 137: 11598–11601
45 Zhang L, Wu YW, Brunsveld L. *Angew Chem Int Ed*, 2007, 46: 1798–1802
46 Dang DT, Nguyen HD, Merkx M, Brunsveld L. *Angew Chem Int Ed*, 2013, 52: 2915–2919
47 Uhlenheuer DA, Young JF, Nguyen HD, Scheepstra M, Brunsveld L. *Chem Commun*, 2011, 47: 6798–6800
48 Nguyen HD, Dang DT, van Dongen JLJ, Brunsveld L. *Angew Chem Int Ed*, 2010, 122: 907–910
49 Uhlenheuer DA, Wasserberg D, Nguyen H, Zhang L, Blum C, Subramaniam V, Brunsveld L. *Chem Eur J*, 2009, 15: 8779–8790
50 Dang DT, Schill J, Brunsveld L. *Chem Sci*, 2012, 3: 2679–2684
51 Hou CX, Li JX, Zhao LL, Zhang W, Luo Q, Dong ZY, Xu JY, Liu JQ. *Angew Chem Int Ed*, 2013, 125: 5700-5703
52 Si CY, Li JX, Luo Q, Hou CX, Pan TZ, Li HB, Liu JQ. *Chem Commun*, 2016, 52: 2924–2927
53 Padilla JE, Colovos C, Yeates TO. *Proc Natl Acad Sci USA*, 2001, 98: 2217–2221
54 Lai YT, Cascio D, Yeates TO. *Science*, 2012, 336: 1129
55 Lai YT, Reading E, Hura GL, Tsai KL, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO. *Nat Chem*, 2014, 6: 1065–1071
56 Usui K, Maki T, Ito F, Suenaga A, Kidoaki S, Itoh M, Taiji M, Matsuda T, Hayashizaki Y, Suzuki H. *Protein Sci*, 2009, 18: 960–969
57 Gao X, Yang S, Zhao CC, Ren YH, Wei DZ. *Angew Chem Int Ed*, 2014, 53: 14027–14030
58 Gonen S, DiMaio F, Gonen T, Baker D. *Science*, 2015, 348: 1365–1368
59 King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D. *Science*, 2012, 336: 1171–1174